Inconsistency-Tolerant Semantics Based on (Preferred) Repairs

Camille Bourgaux

CNRS & DI ENS, Paris, France

Reasoning Web Summer School - September 2025

Database or knowledge base (KB): $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$

- $\mathcal{D}$ dataset: set of facts (ground atoms)
- $\mathcal{T}$ logical theory: set of formulas in some language
  - integrity constraint language (database case)
    - denial constraints: $\forall \vec{x}(\beta[\vec{x}] \wedge \epsilon[\vec{x}] \rightarrow \bot)$
      with $\beta[\vec{x}]$ conjunction of relational atoms
      and $\epsilon[\vec{x}]$ conjunction of inequality atoms
    - universal constraints: $\forall \vec{x}(\beta[\vec{x}] \wedge \epsilon[\vec{x}] \rightarrow \bigvee_{i=1}^{k} \eta_i[\vec{x}])$
    - ...
  - ontology language (KB case)
    - description logic
    - Datalog$^{\pm}$ fragment
    - ...

Conjunctive query: $q(\vec{x}) = \exists \vec{y} \varphi(\vec{x}, \vec{y})$ with $\varphi(\vec{x}, \vec{y})$ conjunction of atoms

Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

- Constant a interpreted as $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
  - $a \neq b$ implies $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if unique name assumption is made
  - $a^{\mathcal{I}} = a$ if standard name assumption is made
- Predicate P of arity n interpreted as a set $P^{\mathcal{I}}$ of n-tuples of $\Delta^{\mathcal{I}}$

Settings: Databases and knowledge bases

Semantics

Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

- Constant a interpreted as $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
  - $a \neq b$ implies $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if unique name assumption is made
  - $a^{\mathcal{I}} = a$ if standard name assumption is made
- Predicate P of arity n interpreted as a set $P^{\mathcal{I}}$ of n-tuples of $\Delta^{\mathcal{I}}$

Database: closed world assumption

- Special interpretation $\mathcal{I}_{\mathcal{D}}$: $P(c_1, \ldots, c_n) \in \mathcal{D}$ iff $(c_1, \ldots, c_n) \in P^{\mathcal{I}_{\mathcal{D}}}$
- $\mathcal{K}$ is consistent if $\mathcal{I}_{\mathcal{D}} \models \mathcal{T}$
- $\mathcal{K} \models q(\vec{a})$ if $\mathcal{I}_{\mathcal{D}} \models q(\vec{a})$

Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

- Constant a interpreted as $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
  - $a \neq b$ implies $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if unique name assumption is made
  - $a^{\mathcal{I}} = a$ if standard name assumption is made
- Predicate P of arity n interpreted as a set $P^{\mathcal{I}}$ of n-tuples of $\Delta^{\mathcal{I}}$

Database: closed world assumption

- Special interpretation $\mathcal{I}_{\mathcal{D}}$: $P(c_1, \ldots, c_n) \in \mathcal{D}$ iff $(c_1, \ldots, c_n) \in P^{\mathcal{I}_{\mathcal{D}}}$
- $\mathcal{K}$ is consistent if $\mathcal{I}_{\mathcal{D}} \models \mathcal{T}$
- $\mathcal{K} \models q(\vec{a})$ if $\mathcal{I}_{\mathcal{D}} \models q(\vec{a})$

KB: open world assumption

- Models: interpretations that satisfy all facts in $\mathcal{D}$ and formulas in $\mathcal{T}$
- $\mathcal{K}$ is consistent if it has some model
- $\mathcal{K} \models q(\vec{a})$ if $q(\vec{a})$ holds in every model of $\mathcal{K}$

Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

- Constant a interpreted as $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
    - $a \neq b$ implies $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if unique name assumption is made
    - $a^{\mathcal{I}} = a$ if standard name assumption is made
- Predicate P of arity n interpreted as a set $P^{\mathcal{I}}$ of n-tuples of $\Delta^{\mathcal{I}}$

Database: closed world assumption

- Special interpretation $\mathcal{I}_{\mathcal{D}}$: $P(c_1, \ldots, c_n) \in \mathcal{D}$ iff $(c_1, \ldots, c_n) \in P^{\mathcal{I}_{\mathcal{D}}}$
- $\mathcal{K}$ is consistent if $\mathcal{I}_{\mathcal{D}} \models \mathcal{T}$
- $\mathcal{K} \models q(\vec{a})$ if $\mathcal{I}_{\mathcal{D}} \models q(\vec{a})$

KB: open world assumption

- Models: interpretations that satisfy all facts in $\mathcal{D}$ and formulas in $\mathcal{T}$
- $\mathcal{K}$ is consistent if it has some model
- $\mathcal{K} \models q(\vec{a})$ if $q(\vec{a})$ holds in every model of $\mathcal{K}$

Database versus KB

- $\mathcal{D} = \{A(a)\}$ is inconsistent with constraint $A(x) \rightarrow B(x)$
- $\mathcal{D} = \{A(a)\}$ and ontological axiom $A(x) \rightarrow B(x)$ entail $B(a)$

Example of description logic ontology

$$\mathsf{Cancer} \sqcap \exists \mathsf{primaryTumor}.\mathsf{Lung} \sqsubseteq \mathsf{LungCancer}$$
$$\mathsf{SmallCellCarcinoma} \sqsubseteq \mathsf{Cancer}$$
$$\mathsf{Adenocarcinoma} \sqsubseteq \mathsf{Cancer}$$
$$\mathsf{Adenocarcinoma} \sqcap \mathsf{SmallCellCarcinoma} \sqsubseteq \bot$$
$$(\mathsf{functional}\ \mathsf{primaryTumor})$$
$$\mathsf{Lung} \sqcap \mathsf{Breast} \sqsubseteq \bot$$

Ontology translation in first-order logic OR database integrity constraints

$$\mathsf{Cancer}(x) \land \mathsf{primaryTumor}(x, y) \land \mathsf{Lung}(y) \to \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \to \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \to \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \land \mathsf{SmallCellCarcinoma}(x) \to \bot$$
$$\mathsf{primaryTumor}(x, y) \land \mathsf{primaryTumor}(x, z) \land y \neq z \to \bot$$
$$\mathsf{Lung}(x) \land \mathsf{Breast}(x) \to \bot$$

Settings: Databases and knowledge bases

Languages

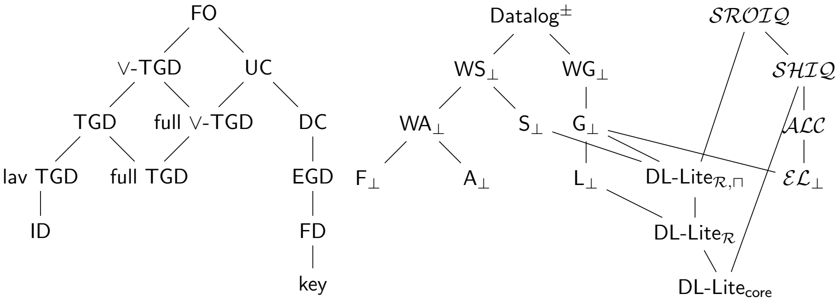


Figure: Hierarchies of database integrity contraint languages (left) [Arming et al., 2016, Fig. 1] and of some ontology languages (right). There is a downward path from $\mathcal{L}_1$ to $\mathcal{L}_2$ if any set of integrity constraints (resp. any ontology) in $\mathcal{L}_2$ can be rewritten into an equivalent set of integrity constraints (resp. ontology) in $\mathcal{L}_1$.

In real world, data often contains errors: human errors, automatic extraction, outdated information...
$\Rightarrow \mathcal{D}$ is likely to be inconsistent with $\mathcal{T}$ ("$\mathcal{T}$-inconsistent")
(focus on the case where $\mathcal{T}$ is consistent and reliable)

Standard semantics when $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ is inconsistent:

- KB case: no model of $\mathcal{K} \Rightarrow$ everything is entailed!
- Database case: query results may be inconsistent with $\mathcal{T}$

It is not always possible to resolve the inconsistencies (lack of information, time, permission...)

Alternative semantics: meaningful answers to queries despite inconsistencies

Handling inconsistent data

Example (KB case)

$$\text{Cancer}(x) \land \text{primaryTumor}(x, y) \land \text{Lung}(y) \rightarrow \text{LungCancer}(x)$$
$$\text{SmallCellCarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \land \text{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\text{primaryTumor}(x, y) \land \text{primaryTumor}(x, z) \rightarrow y = z$$
$$\text{Lung}(x) \land \text{Breast}(x) \rightarrow \bot$$

$\text{hasDisease}(bob, d_1)$

$\text{SmallCellCarcinoma}(d_1)$	$\text{Adenocarcinoma}(d_1)$
$\text{primaryTumor}(d_1, o_1)$	$\text{primaryTumor}(d_1, o_2)$
$\text{Lung}(o_1)$	$\text{Breast}(o_2)$

Handling inconsistent data

Example (KB case)

$$\text{Cancer}(x) \wedge \text{primaryTumor}(x, y) \wedge \text{Lung}(y) \rightarrow \text{LungCancer}(x)$$
$$\text{SmallCellCarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \wedge \text{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\text{primaryTumor}(x, y) \wedge \text{primaryTumor}(x, z) \rightarrow y = z$$
$$\text{Lung}(x) \wedge \text{Breast}(x) \rightarrow \bot$$

$$\text{hasDisease}(bob, d_1)$$

$\text{SmallCellCarcinoma}(d_1)$	$\text{Adenocarcinoma}(d_1)$
$\text{primaryTumor}(d_1, o_1)$	$\text{primaryTumor}(d_1, o_2)$
$\text{Lung}(o_1)$	$\text{Breast}(o_2)$

Handling inconsistent data

Example (KB case)

$$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x,y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\mathsf{primaryTumor}(x,y) \wedge \mathsf{primaryTumor}(x,z) \rightarrow y = z$$
$$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

$\mathcal{K} \models \exists y \mathsf{hasDisease}(x,y) \wedge \mathsf{LungCancer}(y)$ for $x \in \{bob, d_1, d_2, o_1, o_2\}$

$\Rightarrow$ Use inconsistency-tolerant semantics

Ideas? What would you “reasonably” infer about Bob?

Handling inconsistent data

Example (KB case)

$$\mathsf{Cancer}(x) \land \mathsf{primaryTumor}(x,y) \land \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \land \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\mathsf{primaryTumor}(x,y) \land \mathsf{primaryTumor}(x,z) \land y \neq z \rightarrow \bot$$
$$\mathsf{Lung}(x) \land \mathsf{Breast}(x) \rightarrow \bot$$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

$$\text{Cancer}(x) \land \text{primaryTumor}(x,y) \land \text{Lung}(y) \rightarrow \text{LungCancer}(x)$$
$$\text{SmallCellCarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \land \text{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\text{primaryTumor}(x,y) \land \text{primaryTumor}(x,z) \land y \neq z \rightarrow \bot$$
$$\text{Lung}(x) \land \text{Breast}(x) \rightarrow \bot$$

$\text{hasDisease}(bob, d_1)$

~~$\text{SmallCellCarcinoma}(d_1)$~~ $\text{Adenocarcinoma}(d_1)$

~~$\text{primaryTumor}(d_1, o_1)$~~ $\text{primaryTumor}(d_1, o_2)$

$\text{Lung}(o_1)$ $\text{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

Handling inconsistent data

Example (KB case)

$$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x,y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\mathsf{primaryTumor}(x,y) \wedge \mathsf{primaryTumor}(x,z) \wedge y \neq z \rightarrow \bot$$
$$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ ~~$\mathsf{Adenocarcinoma}(d_1)$~~

~~$\mathsf{primaryTumor}(d_1, o_1)$~~ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x,y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$
$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$
$\mathsf{primaryTumor}(x,y) \wedge \mathsf{primaryTumor}(x,z) \wedge y \neq z \rightarrow \bot$
$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$

$\mathsf{hasDisease}(bob, d_1)$

~~$\mathsf{SmallCellCarcinoma}(d_1)$~~ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ ~~$\mathsf{primaryTumor}(d_1, o_2)$~~

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

Handling inconsistent data

Example (KB case)

$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x, y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$
$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$
$\mathsf{primaryTumor}(x, y) \wedge \mathsf{primaryTumor}(x, z) \wedge y \neq z \rightarrow \bot$
$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ ~~$\mathsf{Adenocarcinoma}(d_1)$~~

$\mathsf{primaryTumor}(d_1, o_1)$ ~~$\mathsf{primaryTumor}(d_1, o_2)$~~

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

Handling inconsistent data

Example (KB case)

$$\text{Cancer}(x) \wedge \text{primaryTumor}(x, y) \wedge \text{Lung}(y) \rightarrow \text{LungCancer}(x)$$
$$\text{SmallCellCarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \wedge \text{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\text{primaryTumor}(x, y) \wedge \text{primaryTumor}(x, z) \wedge y \neq z \rightarrow \bot$$
$$\text{Lung}(x) \wedge \text{Breast}(x) \rightarrow \bot$$

hasDisease(bob, d_1)

~~SmallCellCarcinoma(d_1)~~ Adenocarcinoma(d_1)

primaryTumor(d_1, o_1) primaryTumor(d_1, o_2)

~~Lung(o_1)~~ Breast(o_2)

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

Handling inconsistent data

Example (KB case)

$\text{Cancer}(x) \land \text{primaryTumor}(x, y) \land \text{Lung}(y) \rightarrow \text{LungCancer}(x)$
$\text{SmallCellCarcinoma}(x) \rightarrow \text{Cancer}(x)$
$\text{Adenocarcinoma}(x) \rightarrow \text{Cancer}(x)$
$\text{Adenocarcinoma}(x) \land \text{SmallCellCarcinoma}(x) \rightarrow \bot$
$\text{primaryTumor}(x, y) \land \text{primaryTumor}(x, z) \land y \neq z \rightarrow \bot$
$\text{Lung}(x) \land \text{Breast}(x) \rightarrow \bot$

$\text{hasDisease}(bob, d_1)$

$\text{SmallCellCarcinoma}(d_1)$ ~~$\text{Adenocarcinoma}(d_1)$~~

$\text{primaryTumor}(d_1, o_1)$ $\text{primaryTumor}(d_1, o_2)$

~~$\text{Lung}(o_1)$~~ $\text{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

Handling inconsistent data

Example (KB case)

$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x, y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$
$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$
$\mathsf{primaryTumor}(x, y) \wedge \mathsf{primaryTumor}(x, z) \wedge y \neq z \rightarrow \bot$
$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$

$\mathsf{hasDisease}(bob, d_1)$

~~$\mathsf{SmallCellCarcinoma}(d_1)$~~ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ ~~$\mathsf{Breast}(o_2)$~~

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

Example (KB case)

$$\text{Cancer}(x) \wedge \text{primaryTumor}(x, y) \wedge \text{Lung}(y) \rightarrow \text{LungCancer}(x)$$
$$\text{SmallCellCarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \rightarrow \text{Cancer}(x)$$
$$\text{Adenocarcinoma}(x) \wedge \text{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\text{primaryTumor}(x, y) \wedge \text{primaryTumor}(x, z) \wedge y \neq z \rightarrow \bot$$
$$\text{Lung}(x) \wedge \text{Breast}(x) \rightarrow \bot$$

$\text{hasDisease}(bob, d_1)$

$\text{SmallCellCarcinoma}(d_1)$ ~~$\text{Adenocarcinoma}(d_1)$~~

$\text{primaryTumor}(d_1, o_1)$ $\text{primaryTumor}(d_1, o_2)$

$\text{Lung}(o_1)$ ~~$\text{Breast}(o_2)$~~

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent

$$\mathsf{Cancer}(x) \land \mathsf{primaryTumor}(x,y) \land \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \land \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\mathsf{primaryTumor}(x,y) \land \mathsf{primaryTumor}(x,z) \land y \neq z \rightarrow \bot$$
$$\mathsf{Lung}(x) \land \mathsf{Breast}(x) \rightarrow \bot$$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent
- CQA semantics: queries that hold in every repair

$\exists y\, \mathsf{hasDisease}(bob, y) \land \mathsf{Cancer}(y)$ plausible/likely

Handling inconsistent data

Example (KB case)

$$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x,y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\mathsf{primaryTumor}(x,y) \wedge \mathsf{primaryTumor}(x,z) \wedge y \neq z \rightarrow \bot$$
$$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent
- Brave semantics: queries that hold in some repair

$\exists y\, \mathsf{hasDisease}(bob, y) \wedge \mathsf{LungCancer}(y)$ possible

$$\mathsf{Cancer}(x) \wedge \mathsf{primaryTumor}(x,y) \wedge \mathsf{Lung}(y) \rightarrow \mathsf{LungCancer}(x)$$
$$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$$
$$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$$
$$\mathsf{primaryTumor}(x,y) \wedge \mathsf{primaryTumor}(x,z) \wedge y \neq z \rightarrow \bot$$
$$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1)$ $\mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1)$ $\mathsf{primaryTumor}(d_1, o_2)$

$\mathsf{Lung}(o_1)$ $\mathsf{Breast}(o_2)$

- (Subset) repair: inclusion-maximal $\mathcal{R} \subseteq \mathcal{D}$ such that $\mathcal{R}$ is $\mathcal{T}$-consistent
- Intersection semantics: queries that hold in the intersection of all repairs

$\exists y\, \mathsf{hasDisease}(bob, y)$ surest

Dataset repairs: Definition

- Subset repair ($\subseteq$-repair): inclusion-maximal $\mathcal{T}$-consistent $\mathcal{R} \subseteq \mathcal{D}$
- Superset repair ($\supseteq$-repair): inclusion-minimal $\mathcal{T}$-consistent $\mathcal{R} \supseteq \mathcal{D}$
- Symmetric difference repair (Δ-repair): $\mathcal{T}$-consistent $\mathcal{R}$ such that there is no $\mathcal{T}$-consistent $\mathcal{R}'$ with $\mathcal{R}'\Delta\mathcal{D} \subsetneq \mathcal{R}\Delta\mathcal{D}$

Notation: $S\text{-}Rep_x(\mathcal{K}) = S\text{-}Rep_x(\mathcal{D}, \mathcal{T})$: set of all x-repairs of $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$

Dataset repairs: Definition

- Subset repair ($\subseteq$-repair): inclusion-maximal $\mathcal{T}$-consistent $\mathcal{R} \subseteq \mathcal{D}$
- Superset repair ($\supseteq$-repair): inclusion-minimal $\mathcal{T}$-consistent $\mathcal{R} \supseteq \mathcal{D}$
- Symmetric difference repair (Δ-repair): $\mathcal{T}$-consistent $\mathcal{R}$ such that there is no $\mathcal{T}$-consistent $\mathcal{R}'$ with $\mathcal{R}'\Delta\mathcal{D} \subsetneq \mathcal{R}\Delta\mathcal{D}$

Notation: $S\text{-}Rep_x(\mathcal{K}) = S\text{-}Rep_x(\mathcal{D},\mathcal{T})$: set of all x-repairs of $\mathcal{K} = \langle\mathcal{D},\mathcal{T}\rangle$

$\mathsf{primaryTumor}(x, y) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \wedge \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$

$\mathsf{primaryTumor}(d_1, o_1)$
$\mathsf{SmallCellCarcinoma}(d_1)$ $\mathsf{Adenocarcinoma}(d_1)$

$$S\text{-}Rep_\Delta(\mathcal{K}) = \{\emptyset, \{\mathsf{primaryTumor}(d_1, o_1), \mathsf{SmallCellCarcinoma}(d_1), \mathsf{Cancer}(d_1)\}, \{\mathsf{primaryTumor}(d_1, o_1), \mathsf{Adenocarcinoma}(d_1), \mathsf{Cancer}(d_1)\}\}$$

Dataset repairs: Definition

- Subset repair ($\subseteq$-repair): inclusion-maximal $\mathcal{T}$-consistent $\mathcal{R} \subseteq \mathcal{D}$
- Superset repair ($\supseteq$-repair): inclusion-minimal $\mathcal{T}$-consistent $\mathcal{R} \supseteq \mathcal{D}$
- Symmetric difference repair (Δ-repair): $\mathcal{T}$-consistent $\mathcal{R}$ such that there is no $\mathcal{T}$-consistent $\mathcal{R}'$ with $\mathcal{R}' \Delta \mathcal{D} \subsetneq \mathcal{R} \Delta \mathcal{D}$

Notation: $S\text{-}Rep_x(\mathcal{K}) = S\text{-}Rep_x(\mathcal{D}, \mathcal{T})$: set of all x-repairs of $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$

$\mathsf{primaryTumor}(x, y) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{SmallCellCarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \rightarrow \mathsf{Cancer}(x)$
$\mathsf{Adenocarcinoma}(x) \land \mathsf{SmallCellCarcinoma}(x) \rightarrow \bot$

$\mathsf{primaryTumor}(d_1, o_1)$
$\mathsf{SmallCellCarcinoma}(d_1)$ $\quad$ $\mathsf{Adenocarcinoma}(d_1)$

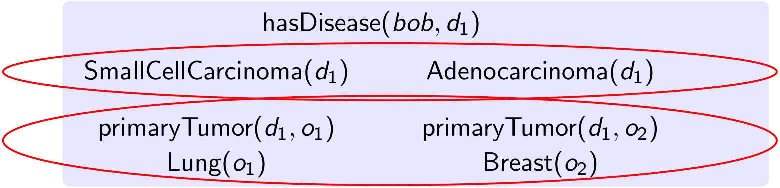
$$S\text{-}Rep_\Delta(\mathcal{K}) = \{\emptyset, \{\mathsf{primaryTumor}(d_1, o_1), \mathsf{SmallCellCarcinoma}(d_1), \mathsf{Cancer}(d_1)\}, \{\mathsf{primaryTumor}(d_1, o_1), \mathsf{Adenocarcinoma}(d_1), \mathsf{Cancer}(d_1)\}\}$$

$\mathcal{K}$ inconsistent KB or database with denial constraints $\Rightarrow S\text{-}Rep_\supseteq(\mathcal{K}) = \emptyset$ and $S\text{-}Rep_\Delta(\mathcal{K}) = S\text{-}Rep_\subseteq(\mathcal{K}) = S\text{-}Rep(\mathcal{K})$

Dataset repairs: Characterization via conflict hypergraph

Case of a KB or database with denial constraints

Conflict: inclusion-minimal $\mathcal{T}$-inconsistent $\mathcal{C} \subseteq \mathcal{D}$
$\Rightarrow$ Conflict hypergraph $\mathcal{G}$: vertices $= \mathcal{D}$, edges $=$ conflicts of $\mathcal{K}$



$\mathcal{R} \in$ *S-Rep*$_{\subseteq}(\mathcal{K})$ iff $\mathcal{R}$ is a maximal independent set of $\mathcal{G}$

Dataset repairs: Characterization via conflict hypergraph

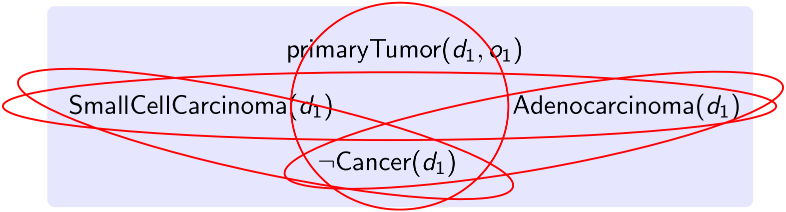
Case of a database with universal constraints: conflicts may contain absent facts

Relevant facts: $Facts_{\mathcal{D}}^{\mathcal{T}} = \mathcal{D} \cup \{P(c_1, \dots, c_n) \mid P \text{ occurs in } \mathcal{T}, c_1, \dots, c_n \text{ occur in } \mathcal{D}\}$

Literals of $\mathcal{D}$: $Lits_{\mathcal{D}}^{\mathcal{T}} = \mathcal{D} \cup \{\neg\alpha \mid \alpha \in Facts_{\mathcal{D}}^{\mathcal{T}} \setminus \mathcal{D}\}$

Conflict: inclusion-minimal $\mathcal{C} \subseteq Lits_{\mathcal{D}}^{\mathcal{T}}$ such that $\mathcal{I} \models \mathcal{C}$ implies $\mathcal{I} \not\models \mathcal{T}$

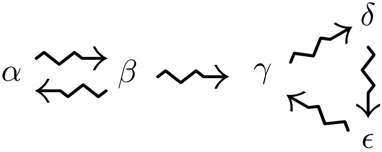
$\Rightarrow$ Conflict hypergraph $\mathcal{G}$: vertices $= Lits_{\mathcal{D}}^{\mathcal{T}}$, edges $=$ conflicts of $\mathcal{K}$



$\mathcal{R} \in S\text{-}Rep_{\Delta}(\mathcal{K})$ iff $Int_{\mathcal{D}}(\mathcal{R})$ is a maximal independent set of $\mathcal{G}$

where $Int_{\mathcal{D}}(\mathcal{R}) = (\mathcal{R} \cap \mathcal{D}) \cup \{\neg\alpha \mid \alpha \in Facts_{\mathcal{D}}^{\mathcal{T}} \setminus (\mathcal{R} \cup \mathcal{D})\}$ is the set of literals upon which $\mathcal{R}$ and $\mathcal{D}$ agree

Abstract argumentation: well-known framework to deal with contradictory information in AI



An (abstract) argumentation framework (AF) is a pair $(Args, \rightsquigarrow)$ where

- *Args* is a finite set of arguments
- $\rightsquigarrow \subseteq Args \times Args$ is the attack relation: α attacks β if $\alpha \rightsquigarrow \beta$

$+$ variant of AF with collective attacks: set-based AF (SETAF)

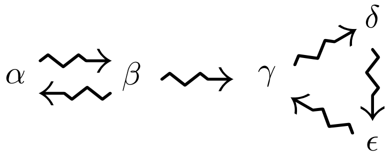
- collective attacks $S \rightsquigarrow \alpha$ with S finite set of arguments

Semantics based on extensions (sets of arguments that represent coherent points of view) $+$ inference mechanism (skeptical or credulous)

Dataset repairs: Connections with abstract argumentation

Several different notions of extension, in particular:

- Naïve extension: $\subseteq$-maximal conflict-free set of arguments
- Preferred extension: $\subseteq$-maximal conflict-free self-defending set (i.e., attacks all arguments that attack some of its arguments)
- Stable extension: conflict-free set attacking all excluded arguments



Naïve: $\{\alpha, \gamma\}$, $\{\alpha, \delta\}$, $\{\alpha, \epsilon\}$, $\{\beta, \delta\}$, $\{\beta, \epsilon\}$

Preferred: $\{\alpha\}$, $\{\beta, \delta\}$

Stable: $\{\beta, \delta\}$

Stable extensions are also preferred extensions

Coherent (SET)AF: stable and preferred extensions coincide

Dataset repairs: Connections with abstract argumentation

Translation of a KB (or database with denial constraints) $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ without self-conflicting fact into a SETAF $F_{\mathcal{K}}$

- Use $\mathcal{D}$ as the arguments
- Define attacks by $\mathcal{C} \setminus \{\alpha\} \rightsquigarrow \alpha$ for every conflict $\mathcal{C}$ and $\alpha \in \mathcal{C}$

$\mathcal{R} \in S\text{-}Rep(\mathcal{K})$ iff $\mathcal{R}$ is a naïve/preferred/stable extension of $F_{\mathcal{K}}$

Dataset repairs: Connections with abstract argumentation

Translation of a KB (or database with denial constraints) $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ without self-conflicting fact into a SETAF $F_{\mathcal{K}}$

- Use $\mathcal{D}$ as the arguments
- Define attacks by $\mathcal{C} \setminus \{\alpha\} \rightsquigarrow \alpha$ for every conflict $\mathcal{C}$ and $\alpha \in \mathcal{C}$

$\mathcal{R} \in S\text{-}Rep(\mathcal{K})$ iff $\mathcal{R}$ is a naïve/preferred/stable extension of $F_{\mathcal{K}}$

Possible to adapt this translation to databases with universal constraints by considering literals instead of facts

Dataset repairs: Connections with abstract argumentation

Translation of a KB (or database with denial constraints) $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ without self-conflicting fact into a SETAF $F_{\mathcal{K}}$

- Use $\mathcal{D}$ as the arguments
- Define attacks by $\mathcal{C} \setminus \{\alpha\} \rightsquigarrow \alpha$ for every conflict $\mathcal{C}$ and $\alpha \in \mathcal{C}$

$\mathcal{R} \in S\text{-}Rep(\mathcal{K})$ iff $\mathcal{R}$ is a naïve/preferred/stable extension of $F_{\mathcal{K}}$

Possible to adapt this translation to databases with universal constraints by considering literals instead of facts

For integrity constraints with existential quantifier in the head, conflicts are not defined but a connection between $\subseteq$-repairs of databases with dependencies of the form $P(\vec{x}) \rightarrow \exists \vec{y} Q[\vec{x}, \vec{y}]$ and $P(\vec{x}) \wedge P(\vec{y}) \wedge \bigwedge_{i \in I} x_i = y_i \rightarrow \bigwedge_{j \in J} x_j = y_j$ and AF extensions has been shown [Mahmood et al., 2024]

Repair-based inconsistency-tolerant semantics

CQA, intersection and brave semantics

- $\vec{a}$ is an answer to $q(\vec{x})$ over $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ under CQA semantics iff $\langle \mathcal{R}, \mathcal{T} \rangle \models q(\vec{a})$ for every $\mathcal{R} \in \textit{S-Rep}(\mathcal{K})$
- $\vec{a}$ is an answer to $q(\vec{x})$ over $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ under intersection semantics iff $\langle \mathcal{R}_\cap, \mathcal{T} \rangle \models q(\vec{a})$ where $\mathcal{R}_\cap = \bigcap_{\mathcal{R} \in \textit{S-Rep}(\mathcal{K})} \mathcal{R}$
- $\vec{a}$ is an answer to $q(\vec{x})$ over $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ under brave semantics iff $\langle \mathcal{R}, \mathcal{T} \rangle \models q(\vec{a})$ for some $\mathcal{R} \in \textit{S-Rep}(\mathcal{K})$

[Arenas et al., 1999, Lembo et al., 2010, Bienvenu and Rosati, 2013]

Repair-based inconsistency-tolerant semantics

CQA, intersection and brave semantics

- $\vec{a}$ is an answer to $q(\vec{x})$ over $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ under CQA semantics iff $\langle \mathcal{R}, \mathcal{T} \rangle \models q(\vec{a})$ for every $\mathcal{R} \in S\text{-}Rep(\mathcal{K})$
- $\vec{a}$ is an answer to $q(\vec{x})$ over $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ under intersection semantics iff $\langle \mathcal{R}_\cap, \mathcal{T} \rangle \models q(\vec{a})$ where $\mathcal{R}_\cap = \bigcap_{\mathcal{R} \in S\text{-}Rep(\mathcal{K})} \mathcal{R}$
- $\vec{a}$ is an answer to $q(\vec{x})$ over $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$ under brave semantics iff $\langle \mathcal{R}, \mathcal{T} \rangle \models q(\vec{a})$ for some $\mathcal{R} \in S\text{-}Rep(\mathcal{K})$

[Arenas et al., 1999, Lembo et al., 2010, Bienvenu and Rosati, 2013]

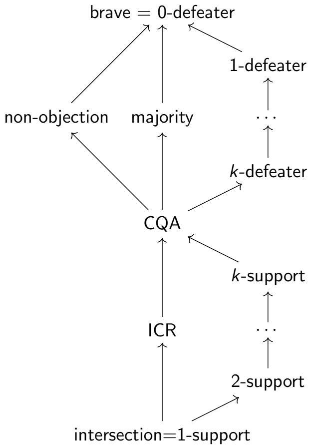
- CQA is the most well-known and accepted semantics
- CQA is usually intractable (coNP-complete in data complexity even for very basic ontology/constraint languages)
- Intersection and brave: under- and over- approximations of CQA:

$$\text{intersection} \rightarrow \text{CQA} \rightarrow \text{brave}$$

- Intersection and brave are tractable for denial constraints and some simple ontology languages

Repair-based inconsistency-tolerant semantics

Overview of other repair-based semantics



	Semantics with the property
Consistent Support	all
Consistent Results	intersection, ICR, k-support, CQA, non-objection
Unique Base	intersection, ICR

- Consistent Support: for every $\langle \mathcal{D}, \mathcal{T} \rangle$, $q(\vec{x})$ and $\vec{a}$, if $\langle \mathcal{D}, \mathcal{T} \rangle \models_{\mathsf{Sem}} q(\vec{a})$, then there exists a $\mathcal{T}$-consistent subset $\mathcal{S}$ of $\mathcal{D}$ such that $\langle \mathcal{S}, \mathcal{T} \rangle \models q(\vec{a})$.
- Consistent Results: for every $\langle \mathcal{D}, \mathcal{T} \rangle$, there exists a model $\mathcal{I}$ of $\mathcal{T}$ such that for every $q(\vec{x})$ and $\vec{a}$, $\langle \mathcal{D}, \mathcal{T} \rangle \models_{\mathsf{Sem}} q(\vec{a})$ implies $\mathcal{I} \models q(\vec{a})$.
- Unique Base: for every $\langle \mathcal{D}, \mathcal{T} \rangle$, there exists a $\mathcal{T}$-consistent dataset $\mathcal{D}'$ such that for every $q(\vec{x})$ and $\vec{a}$, $\langle \mathcal{D}, \mathcal{T} \rangle \models_{\mathsf{Sem}} q(\vec{a})$ iff $\langle \mathcal{D}', \mathcal{T} \rangle \models q(\vec{a})$.

Figure: (left) Relationships between repair-based semantics (adapted from [Bienvenu, 2020]): Sem $\rightarrow$ Sem$'$ means that $\mathcal{K} \models_{\mathsf{Sem}} q(\vec{a})$ implies $\mathcal{K} \models_{\mathsf{Sem}'} q(\vec{a})$. (right) Properties of repair-based semantics (adapted from [Bienvenu, 2020]).

Preferred repairs

In many scenarios, define preferred repairs based on some preference information

- Relative or absolute reliability of facts
- Preference rules
- ...

Preferred repairs

In many scenarios, define preferred repairs based on some preference information

- Relative or absolute reliability of facts
- Preference rules
- ...

Impact of using preferred repairs on repair-based semantics

- More answers hold under CQA/intersection
- Less answers hold under brave
- Relationships between semantics are preserved

hasDisease(bob, d_1)

SmallCellCarcinoma(d_1) Adenocarcinoma(d_1)

primaryTumor(d_1, o_1) primaryTumor(d_1, o_2)
Lung(o_1) Breast(o_2)

Assume two preferred repairs below, which consequences ?

{hasDisease(bob, d_1), SmallCellCarcinoma(d_1), primaryTumor(d_1, o_1), Lung(o_1)}

{hasDisease(bob, d_1), Adenocarcinoma(d_1), primaryTumor(d_1, o_1), Lung(o_1)}

Preferred repairs based on a preorder over datasets

Generalization of the definition of dataset repairs

Strictly $\subseteq$-monotone preorder $\preceq$: reflexive and transitive binary relation over datasets such that $\mathcal{B} \subset \mathcal{B}'$ implies $\mathcal{B} \prec \mathcal{B}'$

- $\preceq$-optimal $\subseteq$-repair: $\mathcal{T}$-consistent dataset $\mathcal{R} \subseteq \mathcal{D}$ such that there is no $\mathcal{T}$-consistent $\mathcal{R}' \subseteq \mathcal{D}$ such that $\mathcal{R} \prec \mathcal{R}'$
- $\preceq$-optimal Δ-repair: $\mathcal{T}$-consistent dataset $\mathcal{R}$ such that there is no $\mathcal{T}$-consistent $\mathcal{R}'$ such that $\mathcal{R}'\Delta\mathcal{D} \prec \mathcal{R}\Delta\mathcal{D}$

$\preceq$ strictly $\subseteq$-monotone implies that $\preceq$-optimal $\subseteq$-repairs are indeed $\subseteq$-repairs, and $\preceq$-optimal Δ-repairs are Δ-repairs

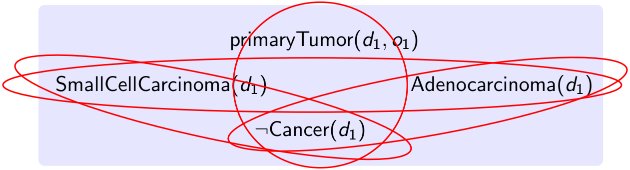
Preferred repairs based on a preorder over datasets

$\leq$-optimal repairs

Cardinality-based repairs: [Lopatenko and Bertossi, 2007]

$$\mathcal{B} \leq \mathcal{B}' \text{ iff } |\mathcal{B}| \leq |\mathcal{B}'|$$

- Fewest modifications
- Appropriate when all facts/literals have same probability of being erroneous



$$S\text{-}Rep_{\Delta}(\mathcal{K}) = \{\emptyset, \{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\}, \{\text{primaryTumor}(d_1, o_1), \text{Adenocarcinoma}(d_1), \text{Cancer}(d_1)\}\}$$

$\leq$-optimal: $\{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\}$

$\{\text{primaryTumor}(d_1, o_1), \text{Adenocarcinoma}(d_1), \text{Cancer}(d_1)\}$

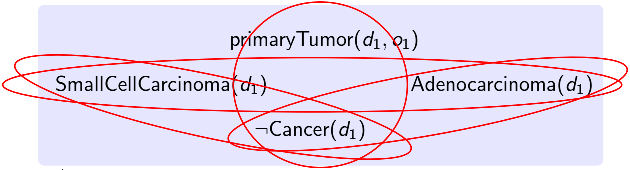
Preferred repairs based on a preorder over datasets

$\leq_w$-optimal repairs

Weight-based repairs: function w assigns weights to facts [Du et al., 2013]

$$\mathcal{B} \leq_w \mathcal{B}' \text{ iff } \Sigma_{\alpha \in \mathcal{B}} w(\alpha) \leq \Sigma_{\alpha \in \mathcal{B}'} w(\alpha)$$

- Model the reliability of facts of $\mathcal{D}$: the higher weight, the more reliable
- Δ-repair case: w assigns weights to all possible facts
  - example: same weight to all facts that do not belong to $\mathcal{D}$



$$S\text{-}Rep_{\Delta}(\mathcal{K}) = \{\emptyset, \{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\}, \{\text{primaryTumor}(d_1, o_1), \text{Adenocarcinoma}(d_1), \text{Cancer}(d_1)\}\}$$

Let $w(\text{Cancer}) = 5$, $w(\text{SmallCell}) = 4$ and $w(\text{primary}) = w(\text{Adeno}) = 1$

$\leq_w$-optimal: $\emptyset$

$\{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\}$

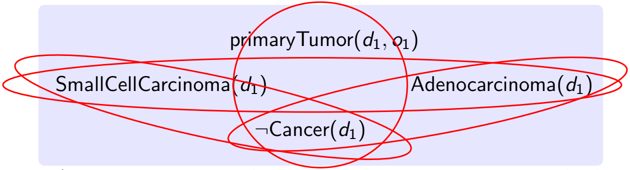
Two kinds of repairs based on priority levels

Prioritization $P = \langle \mathcal{P}_1, \ldots, \mathcal{P}_n \rangle$: disjoint datasets such that $\mathcal{D} = \bigcup_{i=1}^{n} \mathcal{P}_i$

- $\mathcal{P}_1$: most reliable, $\mathcal{P}_n$: least reliable
- Facts coming from different sources, part of the dataset already validated versus recent additions, relative reliability of predicates...
- Best suited when there is a significant difference in the perceived reliability
- Δ-repair case: prioritization of literals ($Lits^{\mathcal{T}}_{\mathcal{D}}$)

Prioritized set inclusion: [Bienvenu et al., 2014]

$$\mathcal{B} \subseteq_P \mathcal{B}' \text{ iff either } \mathcal{B} \cap \mathcal{P}_i = \mathcal{B}' \cap \mathcal{P}_i \text{ for every } 1 \leq i \leq n,$$
$$\text{or there is some } 1 \leq i \leq n \text{ such that}$$
$$\mathcal{B} \cap \mathcal{P}_i \subsetneq \mathcal{B}' \cap \mathcal{P}_i \text{ and } \mathcal{B} \cap \mathcal{P}_j = \mathcal{B}' \cap \mathcal{P}_j \text{ for } 1 \leq j < i$$



$S\text{-}Rep_\Delta(\mathcal{K}) = \{\emptyset, \{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\},$
$\{\text{primaryTumor}(d_1, o_1), \text{Adenocarcinoma}(d_1), \text{Cancer}(d_1)\}\}$

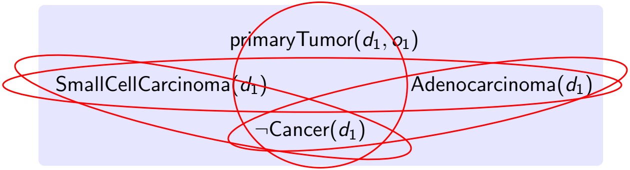
Let $\mathcal{P}_1 = \{\text{SmallCell}, \neg\text{Cancer}\}$, $\mathcal{P}_2 = \{\text{Adeno}, \text{primary}\}$

$\subseteq_P$-optimal: $\emptyset$

$\{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\}$

Prioritized cardinality: [Bienvenu et al., 2014]

$$\mathcal{B} \leq_P \mathcal{B}' \text{ iff either } |\mathcal{B} \cap \mathcal{P}_i| = |\mathcal{B}' \cap \mathcal{P}_i| \text{ for every } 1 \leq i \leq n,$$
$$\text{or there is some } 1 \leq i \leq n \text{ such that}$$
$$|\mathcal{B} \cap \mathcal{P}_i| < |\mathcal{B}' \cap \mathcal{P}_i| \text{ and } |\mathcal{B} \cap \mathcal{P}_j| = |\mathcal{B}' \cap \mathcal{P}_j| \text{ for } 1 \leq j < i$$



$$S\text{-}Rep_\Delta(\mathcal{K}) = \{\emptyset, \{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\},$$
$$\{\text{primaryTumor}(d_1, o_1), \text{Adenocarcinoma}(d_1), \text{Cancer}(d_1)\}\}$$

Let $\mathcal{P}_1 = \{\text{SmallCell}, \neg\text{Cancer}\}$, $\mathcal{P}_2 = \{\text{Adeno}, \text{primary}\}$

$\leq_P$-optimal: $\{\text{primaryTumor}(d_1, o_1), \text{SmallCellCarcinoma}(d_1), \text{Cancer}(d_1)\}$

Properties and relationships

- Weight-based repairs generalize cardinality-based repairs
    - let w assign the same weight to every fact
- Weight-based repairs generalize repairs based on prioritized cardinality
    - let $u = (\max_{i=1}^{n} |\mathcal{P}_i|) + 1$ and $w(\alpha) = u^{n-i}$ for every $\alpha \in \mathcal{P}_i$
- If $P = \langle \mathcal{P}_1 \rangle$, then $\subseteq_P$-optimal = standard and $\leq_P$-optimal = $\leq$-optimal

Optimal repairs based on a priority relation

KB or denial constraints case

When information about relative reliability of facts is available, define priorities between conflicting facts

Examples of possible preferences [Bienvenu et al., 2025]

- Prefer more recent (updated) or older (curated) facts

Fact	Date
primaryTumor(d_1, o_1)	08.10.2023
primaryTumor(d_1, o_2)	05.22.2023

most recent fact gives the last, revised, diagnosis

$\Rightarrow$ primaryTumor(d_1, o_1) $\succ$ primaryTumor(d_1, o_2)

Optimal repairs based on a priority relation

KB or denial constraints case

When information about relative reliability of facts is available, define priorities between conflicting facts

Examples of possible preferences [Bienvenu et al., 2025]

- Prefer more recent (updated) or older (curated) facts
- Prefer facts that come from some source (process, user...)

Fact	Source
Adenocarcinoma(d_1)	X-ray report
SmallCellCarcinoma(d_1)	biopsy report

the second diagnostic method is more reliable

$\Rightarrow$ SmallCellCarcinoma(d_1) $\succ$ Adenocarcinoma(d_1)

Optimal repairs based on a priority relation

KB or denial constraints case

When information about relative reliability of facts is available, define priorities between conflicting facts

Examples of possible preferences [Bienvenu et al., 2025]

- Prefer more recent (updated) or older (curated) facts
- Prefer facts that come from some source (process, user...)
- Take into account presence or absence of other facts in the dataset

  hasDisease(bob, d_1),
  primaryTumor(d_1, o_1), Lung(o_1),
  primaryTumor(d_1, o_2), Breast(o_2),
  gotSurgery(bob, s), BronchialDebridement(s)

  the dataset indicates that the patient got a surgery common in the case of lung cancer but nothing about a breast cancer treatment
  $\Rightarrow$ primaryTumor(d_1, o_1), Lung(o_1) $\succ$ primaryTumor(d_1, o_2), Breast(o_2)

Optimal repairs based on a priority relation

KB or denial constraints case

When information about relative reliability of facts is available, define priorities between conflicting facts

Examples of possible preferences [Bienvenu et al., 2025]

- Prefer more recent (updated) or older (curated) facts
- Prefer facts that come from some source (process, user...)
- Take into account presence or absence of other facts in the dataset
- ...

hasDisease(bob, d_1)

SmallCellCarcinoma(d_1) $\succ$ Adenocarcinoma(d_1)

primaryTumor(d_1, o_1) $\succ$ primaryTumor(d_1, o_2)

Lung(o_1) $\succ$ Breast(o_2)

Optimal repairs based on a priority relation

KB or denial constraints case

Formally:

- Priority relation $\succ$: acyclic binary relation over $\mathcal{D}$ such that $\alpha \succ \beta$ implies $\{\alpha, \beta\} \subseteq \mathcal{C}$ for some conflict $\mathcal{C}$
- Prioritized KB (or database with denial constraints) $\mathcal{K}_{\succ} = (\mathcal{K}, \succ)$

$\mathsf{hasDisease}(bob, d_1)$

$\mathsf{SmallCellCarcinoma}(d_1) \succ \mathsf{Adenocarcinoma}(d_1)$

$\mathsf{primaryTumor}(d_1, o_1) \succ \mathsf{primaryTumor}(d_1, o_2)$

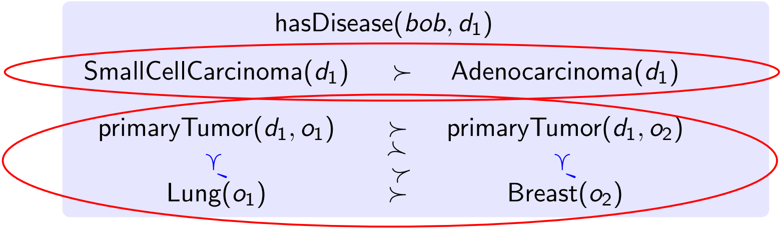
$\mathsf{Lung}(o_1) \succ \mathsf{Breast}(o_2)$

Optimal repairs based on a priority relation

KB or denial constraints case

Formally:

- Priority relation $\succ$: acyclic binary relation over $\mathcal{D}$ such that $\alpha \succ \beta$ implies $\{\alpha, \beta\} \subseteq \mathcal{C}$ for some conflict $\mathcal{C}$
- Prioritized KB (or database with denial constraints) $\mathcal{K}_{\succ} = (\mathcal{K}, \succ)$



- $\succ$ is total if for all $\alpha \neq \beta$ such that $\{\alpha, \beta\} \subseteq \mathcal{C}$ for some conflict $\mathcal{C}$, either $\alpha \succ \beta$ or $\beta \succ \alpha$
- Completion of $\succ$: total priority relation $\succ' \supseteq \succ$
  - example: complete $\succ$ with primaryTumor$(d_1, o_1) \succ'$ Lung(o_1) and primaryTumor$(d_1, o_2) \succ'$ Breast(o_2)

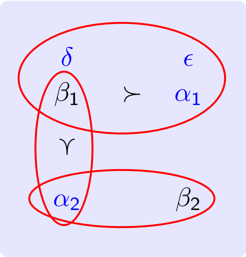
Optimal repairs based on a priority relation

KB or denial constraints case

Three notions of optimal repair [Staworko et al., 2012]

Let $\mathcal{R}$ be a repair of $\mathcal{K}$ ($\mathcal{R} \in$ *S-Rep*$(\mathcal{K})$)

- A Pareto improvement of $\mathcal{R}$ is a $\mathcal{T}$-consistent $\mathcal{B} \subseteq \mathcal{D}$ such that there is $\beta \in \mathcal{B} \setminus \mathcal{R}$ with $\beta \succ \alpha$ for every $\alpha \in \mathcal{R} \setminus \mathcal{B}$
- $\mathcal{R}$ is Pareto-optimal ($\mathcal{R} \in$ *P-Rep*$(\mathcal{K}_\succ)$) if there is no Pareto improvement of $\mathcal{R}$



$\{\alpha_1, \alpha_2, \delta, \epsilon\} \in$ *S-Rep*$(\mathcal{K})$

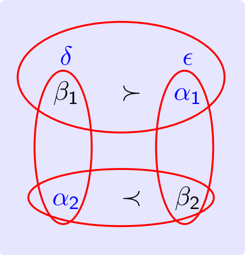
$\{\beta_1, \delta, \epsilon\}$ Pareto improvement

$\Rightarrow \quad \{\alpha_1, \alpha_2, \delta, \epsilon\} \notin$ *P-Rep*$(\mathcal{K}_\succ)$

Three notions of optimal repair [Staworko et al., 2012]

Let $\mathcal{R}$ be a repair of $\mathcal{K}$ ($\mathcal{R} \in$ *S-Rep*$(\mathcal{K})$)

- A global improvement of $\mathcal{R}$ is a $\mathcal{T}$-consistent $\mathcal{B} \subseteq \mathcal{D}$ such that $\mathcal{B} \neq \mathcal{R}$ and for every $\alpha \in \mathcal{R} \setminus \mathcal{B}$, there is $\beta \in \mathcal{B} \setminus \mathcal{R}$ such that $\beta \succ \alpha$
- $\mathcal{R}$ is globally-optimal ($\mathcal{R} \in$ *G-Rep*$(\mathcal{K}_\succ)$) if there is no global improvement of $\mathcal{R}$



$\{\alpha_1, \alpha_2, \delta, \epsilon\} \in$ *P-Rep*$(\mathcal{K}_\succ)$

$\{\beta_1, \beta_2, \delta, \epsilon\}$ global improvement

$\Rightarrow \quad \{\alpha_1, \alpha_2, \delta, \epsilon\} \notin$ *G-Rep*$(\mathcal{K}_\succ)$

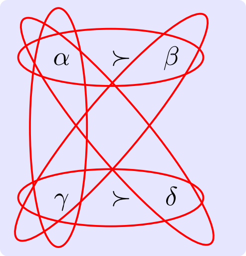
Optimal repairs based on a priority relation

KB or denial constraints case

Three notions of optimal repair [Staworko et al., 2012]

Let $\mathcal{R}$ be a repair of $\mathcal{K}$ ($\mathcal{R} \in$ *S-Rep*$(\mathcal{K})$)

- $\mathcal{R}$ is completion-optimal ($\mathcal{R} \in$ *C-Rep*$(\mathcal{K}_{\succ})$) if $\mathcal{R}$ is globally-optimal w.r.t. some completion $\succ'$ of $\succ$
- Equivalently: obtained by greedily selecting some fact maximal w.r.t. $\succ$ among those not yet considered, and keeping it if still consistent



Subset repairs

$\textit{S-Rep}(\mathcal{K}) = \{ \{\alpha\}, \{\gamma\}, \{\beta, \delta\} \}$

Pareto- and globally-optimal

$\textit{P-Rep}(\mathcal{K}_{\succ}) = \textit{G-Rep}(\mathcal{K}_{\succ}) = \{ \{\alpha\}, \{\gamma\}, \{\beta, \delta\} \}$

Completion-optimal

$\textit{C-Rep}(\mathcal{K}_{\succ}) = \{ \{\alpha\}, \{\gamma\} \}$

Optimal repairs based on a priority relation

KB or denial constraints case

Three notions of optimal repair [Staworko et al., 2012]

$$C\text{-}Rep(\mathcal{K}_{\succ}) \subseteq G\text{-}Rep(\mathcal{K}_{\succ}) \subseteq P\text{-}Rep(\mathcal{K}_{\succ}) \subseteq S\text{-}Rep(\mathcal{K})$$

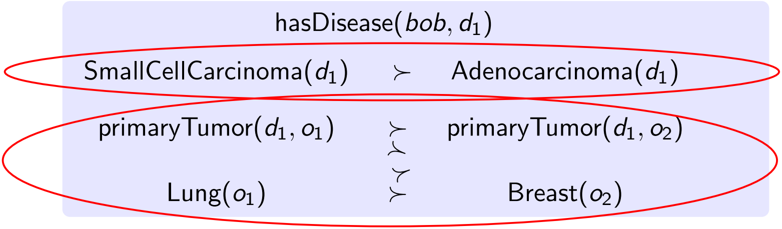
Optimal repairs based on a priority relation

KB or denial constraints case

Three notions of optimal repair [Staworko et al., 2012]

$$C\text{-}Rep(\mathcal{K}_{\succ}) \subseteq G\text{-}Rep(\mathcal{K}_{\succ}) \subseteq P\text{-}Rep(\mathcal{K}_{\succ}) \subseteq S\text{-}Rep(\mathcal{K})$$

If $\succ$ is score-structured (i.e., can be induced by assigning scores to facts and from which we can obtain a prioritization P of $\mathcal{D}$), then $C\text{-}Rep(\mathcal{K}_{\succ}) = G\text{-}Rep(\mathcal{K}_{\succ}) = P\text{-}Rep(\mathcal{K}_{\succ}) = \subseteq_P\text{-}Rep(\mathcal{K})$



$\{$hasDisease(bob, d_1), SmallCellCarcinoma(d_1), primaryTumor(d_1, o_1), Lung(o_1), primaryTumor$(d_1, o_2)\}$

$\{$hasDisease(bob, d_1), SmallCellCarcinoma(d_1), primaryTumor(d_1, o_1), Lung(o_1), Breast$(o_2)\}$

Optimal repairs based on a priority relation

Database with universal constraints case

All definitions and results extend to database with universal constraints by considering literals from $Lits^{\mathcal{T}}_{\mathcal{D}}$ instead of facts from $\mathcal{D}$

[Bienvenu and Bourgaux, 2023]

Preference-based set based argumentation framework (PSETAF)

- Preference relation $\succ$ between arguments
- Refines the attack relation: $S \rightsquigarrow_{\succ} \alpha$ if $S \rightsquigarrow \alpha$ and $\alpha \not\succ \beta$ for every $\beta \in S$

Optimal repairs based on a priority relation

Links with abstract argumentation

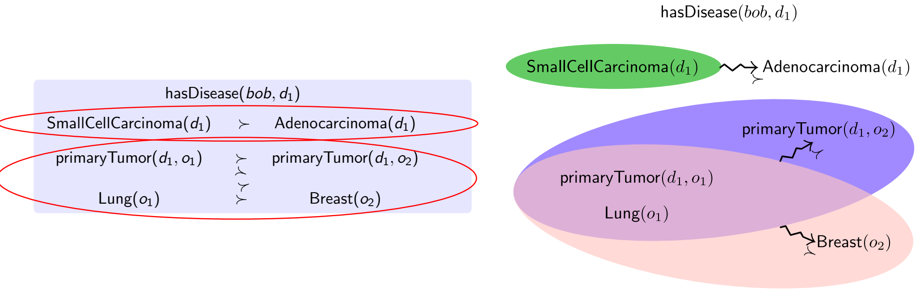
Preference-based set based argumentation framework (PSETAF)

- Preference relation $\succ$ between arguments
- Refines the attack relation: $S \rightsquigarrow_{\succ} \alpha$ if $S \rightsquigarrow \alpha$ and $\alpha \not\succ \beta$ for every $\beta \in S$

Translation of a prioritized KB $\mathcal{K}_{\succ} = (\mathcal{K}, \succ)$ into a PSETAF $F_{\mathcal{K},\succ}$

- Use $\mathcal{D}$ as the arguments
- Use $\succ$ as the preference relation
- Define attacks by $\mathcal{C} \setminus \{\alpha\} \rightsquigarrow \alpha$ for every conflict $\mathcal{C}$ and $\alpha \in \mathcal{C}$

  $\Rightarrow \quad \mathcal{C} \setminus \{\alpha\} \rightsquigarrow_{\succ} \alpha$ if $\alpha \not\succ \beta$ for every $\beta \in \mathcal{C}$



$\mathcal{R}$ is a Pareto-optimal repair of $\mathcal{K}_{\succ}$
iff
$\mathcal{R}$ is a stable extension of $F_{\mathcal{K},\succ}$

If $\succ$ is transitive or if $\mathcal{K}$ has only binary conflicts, then $F_{\mathcal{K},\succ}$ is coherent:
$\mathcal{R}$ is a Pareto-optimal repair of $\mathcal{K}_{\succ}$
iff
$\mathcal{R}$ is a preferred extension of $F_{\mathcal{K},\succ}$

No notion of extension corresponds to globally- or completion-optimal

[Bienvenu and Bourgaux, 2020]

Optimal repairs based on a priority relation

Links with abstract argumentation

The grounded extension of a (PSET)AF is the minimal conflict-free set of arguments that contains all arguments that it defends

- Add all arguments with no incoming attacks
- Iteratively add arguments defended by the selected arguments

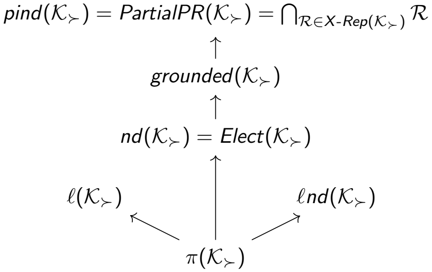
$\Rightarrow$ Grounded semantics for prioritized KB: query grounded extension of $F_{\mathcal{K},\succ}$

- P-complete data complexity for DL-Lite KBs
- Under-approximation of intersection semantics based on Pareto-optimal repairs

"Grounded repair": in the line of research that aims at selecting a single consistent set of facts to query ("unique repair") from a prioritized KB (with $\succ$ often induced by prioritization $P = \langle \mathcal{P}_1, \dots, \mathcal{P}_n \rangle$)

Optimal repairs based on a priority relation

Relationships with "unique repairs"



- Polynomial computation: given the conflict hypergraph and $\succ$, the repair can be computed in polynomial time
  $\Longrightarrow$ holds for all but $pind(\mathcal{K}_\succ)$
- Complete w.r.t. isolated vertices: contains all facts that do not belong to any conflict (i.e., includes $\mathcal{R}_\cap = \bigcap_{\mathcal{R} \in S\text{-}Rep(\mathcal{K})} \mathcal{R}$)
  $\Longrightarrow$ holds for all but $\pi(\mathcal{K}_\succ)$, $\ell(\mathcal{K}_\succ)$
- Sound w.r.t. prioritized intersection: included in the intersection of optimal repairs
  $\Longrightarrow$ holds for all but $\ell(\mathcal{K}_\succ)$, $\ell nd(\mathcal{K}_\succ)$

Figure: (left) Relationships between "unique repairs" when $\succ$ is induced by a prioritization: an arrow $X \rightarrow Y$ means that $X \subseteq Y$ [Benferhat et al., 2015, Belabbes et al., 2021, Bienvenu and Bourgaux, 2020]
(right) Relevant properties [Benferhat et al., 2015, Bienvenu and Bourgaux, 2020]

In the database setting, active integrity constraints state how to resolve constraint violations: high-level similarities with prioritized databases

Example of denial constraint and active denial constraint:

$$\text{Child}(x) \wedge \text{Adult}(x) \rightarrow \bot$$
$$\text{Child}(x) \wedge \text{Adult}(x) \rightarrow \{-\text{Child}(x)\}$$

Example of universal constraint and active universal constraint:

$$\text{Lung}(x) \wedge \neg\text{LeftLg}(x) \wedge \neg\text{RightLg}(x) \rightarrow \bot$$
$$\text{Lung}(x) \wedge \neg\text{LeftLg}(x) \wedge \neg\text{RightLg}(x) \rightarrow \{+\text{LeftLg}(x), +\text{RightLg}(x)\}$$

A ground active integrity constraint (AIC) is a formula of the form

$$\alpha_1 \wedge \cdots \wedge \alpha_n \wedge \neg\beta_1 \wedge \cdots \wedge \neg\beta_m \rightarrow \{A_1, \ldots, A_k\}$$

with update actions A_i of the form $-\alpha_j$ or $+\beta_j$

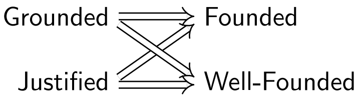
Optimal repairs based on a priority relation

Links with active integrity constraints

Semantics based on repair updates: $\mathcal{U}$ set of update actions such that $\mathcal{D} \circ \mathcal{U}$ is a repair, where $\mathcal{D} \circ \mathcal{U} = \mathcal{D} \setminus \{\alpha \mid -\alpha \in \mathcal{U}\} \cup \{\alpha \mid +\alpha \in \mathcal{U}\}$

Several different notions of repair update, in particular:

- Founded: for every $A \in \mathcal{U}$, there is an AIC r with update action A and $\mathcal{D} \circ \mathcal{U} \setminus \{A\} \not\models r$
- Well-founded: there exists a sequence of actions $A_1, \ldots, A_n$ such that $\mathcal{U} = \{A_1, \ldots, A_n\}$, and for every $1 \leq i \leq n$, there is r_i with update action A_i and $\mathcal{D} \circ \{A_1, \ldots, A_{i-1}\} \not\models r_i$
- Grounded (for normalized AICs: single update action): for every $\mathcal{V} \subsetneq \mathcal{U}$, there is r whose update action is in $\mathcal{U} \setminus \mathcal{V}$ and $\mathcal{D} \circ \mathcal{V} \not\models r$
- Justified...

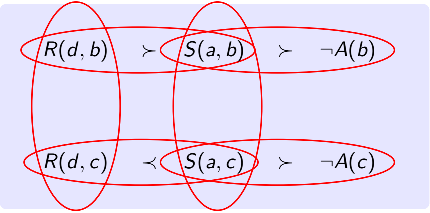


Optimal repairs based on a priority relation

Links with active integrity constraints

Translation of prioritized database (univ. constraints) $\mathcal{K}_\succ$ into ground AICs

- $\eta_\succ^\mathcal{T} = \{r_\mathcal{C} \mid \mathcal{C} \text{ conflict}\}$ where $fix(\alpha) = -\alpha$, $fix(\neg\alpha) = +\alpha$ and
  $r_\mathcal{C} := \bigwedge_{\lambda\in\mathcal{C}} \lambda \rightarrow \{fix(\lambda) \mid \lambda \in \mathcal{C}, \forall\mu \in \mathcal{C}, \lambda \not\succ \mu\}$
- Conflicts fixed by modifying least preferred literals according to $\succ$



$$S(a,b) \wedge S(a,c) \rightarrow \{-S(a,b), -S(a,c)\}$$
$$R(d,b) \wedge R(d,c) \rightarrow \{-R(d,b), -R(d,c)\}$$
$$R(d,b) \wedge S(a,b) \rightarrow \{-S(a,b)\}$$
$$R(d,c) \wedge S(a,c) \rightarrow \{-R(d,c)\}$$
$$S(a,b) \wedge \neg A(b) \rightarrow \{+A(b)\}$$
$$S(a,c) \wedge \neg A(c) \rightarrow \{+A(c)\}$$

For denial constraints: data-independent reduction to non-ground AICs (assuming the priority relation is stored in the database)

Optimal repairs based on a priority relation

Links with active integrity constraints

Pareto ≡ Founded ≡ Grounded ≡ Justified ⇒ Well-Founded

$\mathcal{R} = \mathcal{D} \circ \mathcal{U}$ is a Pareto-optimal repair of $\mathcal{K}_{\succ}$

iff

$\mathcal{U}$ is a founded repair update of $\mathcal{D}$ w.r.t. $\eta_{\succ}^{\mathcal{T}}$

iff

$\mathcal{U}$ is a grounded repair update of $\mathcal{D}$ w.r.t. $\eta_{\succ}^{\mathcal{T}}$

iff

$\mathcal{U}$ is a justified repair update of $\mathcal{D}$ w.r.t. $\eta_{\succ}^{\mathcal{T}}$

[Bienvenu and Bourgaux, 2023]

Preferred repairs based on preference rules

Definition

Define preference rules that must be maximally satisfied by the repairs

- Preference program Π: set of preference rules of form $P[\vec{x}] \succ Q[\vec{y}] \leftarrow \exists \vec{z} \varphi[\vec{z}, \vec{w}]$
  - $\text{Lung}(x) \succ \text{Breast}(x) \leftarrow \exists y \ \text{primaryTumor}(y, x) \land \text{LungCancer}(y)$
- $\mathcal{B}$ satisfies ground preference rule $P(\vec{a}) \succ Q(\vec{b}) \leftarrow \exists \vec{z} \varphi[\vec{z}, \vec{c}]$ if $\langle \mathcal{B}, \mathcal{T} \rangle \models \exists \vec{z} \varphi[\vec{z}, \vec{c}]$ implies that $\langle \mathcal{B}, \mathcal{T} \rangle \not\models Q(\vec{b})$ or $\langle \mathcal{B}, \mathcal{T} \rangle \models P(\vec{a})$
  - $\{\text{primaryTumor}(d, o), \text{LungCancer}(d), \text{Breast}(o)\}$ ✗
  - $\{\text{primaryTumor}(d, o), \text{LungCancer}(d), \text{Lung}(o)\}$ ✓
  - $\{\text{primaryTumor}(d, o), \text{LungCancer}(d)\}$ ✓
  - $\emptyset$ ✓
- For $\trianglelefteq \in \{\subseteq, \leq\}$, repair $\mathcal{R}$ is $\Pi_{\trianglelefteq}$-preferred iff there is no repair $\mathcal{R}'$ such that $GrSat(\mathcal{R}, \mathcal{K}_\Pi) \triangleleft GrSat(\mathcal{R}', \mathcal{K}_\Pi)$, where $GrSat(\mathcal{B}, \mathcal{K}_\Pi)$ denotes the set of ground instances of preference rules satisfied by $\mathcal{B}$

[Calautti et al., 2022]

Preferred repairs based on preference rules

Example

$$\mathsf{Lung}(x) \succ \mathsf{Breast}(x) \leftarrow \exists y\ \mathsf{primaryTumor}(y,x) \wedge \mathsf{LungCancer}(y)$$

$$\mathsf{Breast}(x) \succ \mathsf{Lung}(x) \leftarrow \exists y\ \mathsf{primaryTumor}(y,x) \wedge \mathsf{BreastCancer}(y)$$

$\mathsf{Lung}(x) \wedge \mathsf{Breast}(x) \rightarrow \bot$
$\mathsf{LungCancer}(x) \wedge \mathsf{BreastCancer}(x) \rightarrow \bot$

$\mathsf{primaryTumor}(d,o)$
$\mathsf{Lung}(o)$ $\quad$ $\mathsf{Breast}(o)$
$\mathsf{LungCancer}(d)$ $\quad$ $\mathsf{BreastCancer}(d)$

$$\begin{aligned} \mathit{S\text{-}Rep}(\mathcal{K}) = \{ & \{\mathsf{primaryTumor}(d,o), \mathsf{Lung}(o), \mathsf{LungCancer}(d)\},\\ & \{\mathsf{primaryTumor}(d,o), \mathsf{Breast}(o), \mathsf{LungCancer}(d)\},\\ & \{\mathsf{primaryTumor}(d,o), \mathsf{Lung}(o), \mathsf{BreastCancer}(d)\},\\ & \{\mathsf{primaryTumor}(d,o), \mathsf{Breast}(o), \mathsf{BreastCancer}(d)\}\} \end{aligned}$$

$\Pi_{\subseteq}$-preferred / $\Pi_{\leq}$-preferred: $\{\mathsf{primaryTumor}(d,o), \mathsf{Lung}(o), \mathsf{LungCancer}(d)\}$, $\{\mathsf{primaryTumor}(d,o), \mathsf{Breast}(o), \mathsf{BreastCancer}(d)\}$

Decision problems (parametrized by the kind of repair considered):

- (Preferred) repair checking
  - input: $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$, $\mathcal{R}$
  - output: 'yes' if $\mathcal{R}$ is a (preferred) repair of $\mathcal{K}$, 'no' otherwise
- Boolean conjunctive query entailment under (preferred repair-based) CQA (resp. intersection, brave) semantics
  - input: $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$, q
  - output: 'yes' if q is entailed by $\mathcal{K}$ under (preferred repair-based) CQA (resp. intersection, brave) semantics, 'no' otherwise

Decision problems (parametrized by the kind of repair considered):

- (Preferred) repair checking
  - input: $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$, $\mathcal{R}$
  - output: 'yes' if $\mathcal{R}$ is a (preferred) repair of $\mathcal{K}$, 'no' otherwise
- Boolean conjunctive query entailment under (preferred repair-based) CQA (resp. intersection, brave) semantics
  - input: $\mathcal{K} = \langle \mathcal{D}, \mathcal{T} \rangle$, q
  - output: 'yes' if q is entailed by $\mathcal{K}$ under (preferred repair-based) CQA (resp. intersection, brave) semantics, 'no' otherwise

Complexity measures:

- Combined complexity: in terms of size of whole input
- Data complexity: in terms of size of $\mathcal{D}$ only

Complexity considerations

Standard repair checking

		Data complexity	Combined complexity
DLs $S\text{-}Rep(\mathcal{K})$	DL-Lite$_{\mathcal{R}}$	in L	NL
	DL-Lite$_{\mathcal{R},\sqcap}$	in L	P
	$\mathcal{EL}_\bot$	P	P
	$\mathcal{ALC}$	DP	Exp
	$\mathcal{SHIQ}$	DP	Exp
Datalog$^\pm$ $S\text{-}Rep(\mathcal{K})$	$L_\bot$	in L	PSpace
	$A_\bot$	in L	DExp
	$G_\bot$	P	2Exp
	$S_\bot$	in L	Exp
	$F_\bot$	P	Exp
ICs $S\text{-}Rep(\mathcal{K})$	FD	in L	in L
	DC	in L	DP
ICs $S\text{-}Rep_\Delta(\mathcal{K})$	full TGD	P	DP
	UC	coNP	Π_2^P
	TGD	coNP	Π_3^P

Complexity considerations

Standard repair checking

		Data complexity	Combined complexity
DLs S-$Rep(\mathcal{K})$	DL-Lite$_{\mathcal{R}}$	in L	NL
	DL-Lite$_{\mathcal{R},\sqcap}$	in L	P
	$\mathcal{EL}_{\bot}$	P	P
	$\mathcal{ALC}$	DP	Exp
	$\mathcal{SHIQ}$	DP	Exp
Datalog$^{\pm}$ S-$Rep(\mathcal{K})$	$L_{\bot}$	in L	PSpace
	$A_{\bot}$	in L	DExp
	$G_{\bot}$	P	2Exp
	$S_{\bot}$	in L	Exp
	$F_{\bot}$	P	Exp
ICs S-$Rep(\mathcal{K})$	FD	in L	in L
	DC	in L	DP
ICs S-$Rep_{\Delta}(\mathcal{K})$	full TGD	P	DP
	UC	coNP	Π_2^P
	TGD	coNP	Π_3^P

Main algorithms:

- Non-deterministic algorithm
  - check that $\mathcal{R}$ is $\mathcal{T}$-consistent
  - to show that $\mathcal{R}$ does not minimally differ from $\mathcal{D}$, guess $\mathcal{R}'$ such that $\langle \mathcal{R}', \mathcal{T} \rangle \not\models \bot$ and $\mathcal{R}' \Delta \mathcal{D} \subsetneq \mathcal{R} \Delta \mathcal{D}$
- Deterministic algorithm for $\subseteq$-repairs (KB / denial constraints)
  - check that $\mathcal{R} \subseteq \mathcal{D}$ and $\mathcal{R}$ is $\mathcal{T}$-consistent
  - check that for each $\alpha \in \mathcal{D} \setminus \mathcal{R}$, $\langle \mathcal{R} \cup \{\alpha\}, \mathcal{T} \rangle \models \bot$

Complexity considerations

BCQ entailment under CQA (resp. intersection, brave) semantics

		Data complexity			Combined complexity		
		CQA	Int.	brave	CQA	Int.	brave
DLs $S\text{-}Rep(\mathcal{K})$	DL-Lite$_{\mathcal{R}}$	coNP	in AC^0	in AC^0	Π_2^P	NP	NP
	DL-Lite$_{\mathcal{R},\sqcap}$	coNP	in AC^0	in AC^0	Π_2^P	Θ_2^P	NP
	$\mathcal{EL}_\bot$	coNP	coNP	NP	Π_2^P	Θ_2^P	NP
	$\mathcal{ALC}$	Π_2^P	Π_2^P	Σ_2^P	Exp	Exp	Exp
	$\mathcal{SHIQ}$	Π_2^P	Π_2^P	Σ_2^P	2Exp	2Exp	2Exp
Datalog$^\pm$ $S\text{-}Rep(\mathcal{K})$	$L_\bot$	coNP	in AC^0	in AC^0	PSpace	PSpace	PSpace
	$A_\bot$	coNP	in AC^0	in AC^0	P^{NExp}	P^{NExp}	P^{NExp}
	$G_\bot$	coNP	coNP	NP	2Exp	2Exp	2Exp
	$S_\bot$	coNP	in AC^0	in AC^0	Exp	Exp	Exp
	$F_\bot$	coNP	coNP	NP	Exp	Exp	Exp
ICs $S\text{-}Rep(\mathcal{K})$	FD	coNP	in AC^0	in AC^0	Π_2^P		
	DC	coNP	in AC^0	in AC^0	Π_2^P		
ICs $S\text{-}Rep_\Delta(\mathcal{K})$	full TGD	coNP	coNP	P	Exp		
	UC	Π_2^P	Π_2^P	Σ_2^P	Π_2^{Exp}		
	TGD	undec.			undec.		

Complexity considerations

BCQ entailment under CQA (resp. intersection, brave) semantics

		Data complexity CQA	Data complexity Int.	Data complexity brave	Combined complexity CQA	Combined complexity Int.	Combined complexity brave
DLs S-$Rep(\mathcal{K})$	DL-Lite$_\mathcal{R}$	coNP	in AC^0	in AC^0	Π_2^P	NP	NP
	DL-Lite$_{\mathcal{R},\sqcap}$	coNP	in AC^0	in AC^0	Π_2^P	Θ_2^P	NP
	$\mathcal{EL}_\bot$	coNP	coNP	NP	Π_2^P	Θ_2^P	NP
	$\mathcal{ALC}$	Π_2^P	Π_2^P	Σ_2^P	Exp	Exp	Exp
	$\mathcal{SHIQ}$	Π_2^P	Π_2^P	Σ_2^P	2Exp	2Exp	2Exp
Datalog$^\pm$ S-$Rep(\mathcal{K})$	$L_\bot$	coNP	in AC^0	in AC^0	PSpace	PSpace	PSpace
	$A_\bot$	coNP	in AC^0	in AC^0	P^{NExp}	P^{NExp}	P^{NExp}
	$G_\bot$	coNP	coNP	NP	2Exp	2Exp	2Exp
	$S_\bot$	coNP	in AC^0	in AC^0	Exp	Exp	Exp
	$F_\bot$	coNP	coNP	NP	Exp	Exp	Exp
ICs S-$Rep(\mathcal{K})$	FD	coNP	in AC^0	in AC^0	Π_2^P		
	DC	coNP	in AC^0	in AC^0	Π_2^P		
ICs S-$Rep_\Delta(\mathcal{K})$	full TGD	coNP	coNP	P	Exp		
	UC	Π_2^P	Π_2^P	Σ_2^P	Π_2^{Exp}		
	TGD	undec.			undec.		

Main algorithms:

- Brave/not CQA: guess repair $\mathcal{R}$ such that $\langle \mathcal{R}, \mathcal{T} \rangle \models q$ / $\langle \mathcal{R}, \mathcal{T} \rangle \not\models q$
- Not intersection: guess $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ and repairs $\mathcal{R}_1, \dots, \mathcal{R}_n$ such that $\alpha_i \notin \mathcal{R}_i$ and $\langle \mathcal{D} \setminus \mathcal{B}, \mathcal{T} \rangle \not\models q$
- AC^0 upper bounds via FO rewriting
  $\mathcal{T} = \{A \sqsubseteq B, A \sqsubseteq \neg C\}$, $q(x) = B(x) \Rightarrow q_\cap^\mathcal{T}(x) = B(x) \vee (A(x) \wedge \neg C(x))$

	Repair checking	CQA	Intersection	Brave
$\leq$-optimal	coNP	Θ_2^P	Θ_2^P	Θ_2^P
$\leq_w$-optimal	coNP	Δ_2^P †	Δ_2^P †	Δ_2^P †
$\leq_P$-optimal	coNP	Δ_2^P †	Δ_2^P †	Δ_2^P †
$\subseteq_P$-optimal	in P	coNP	coNP	NP
Pareto-optimal	in P	coNP	coNP	NP
Completion-optimal	in P	coNP	coNP	NP
Globally-optimal	coNP	Π_2^P	Π_2^P	Σ_2^P
$\Pi_\subseteq$-preferred	coNP	Π_2^P	Π_2^P	Σ_2^P
$\Pi_\leq$-preferred	coNP	Θ_2^P	Θ_2^P	Θ_2^P

†: Θ_2^P if there is a data-independent bound on the weights/number of priority levels

Upper bounds: $\subseteq$-repairs and languages with consistency checking/BCQ entailment in P

Lower bounds: DL-Lite$_{\text{core}}$, functional dependencies, or negative constraints

Complexity considerations

Impact of using preferred repairs on data complexity

	Repair checking	CQA	Intersection	Brave
$\leq$-optimal	coNP	Θ_2^P	Θ_2^P	Θ_2^P
$\leq_w$-optimal	coNP	Δ_2^P †	Δ_2^P †	Δ_2^P †
$\leq_P$-optimal	coNP	Δ_2^P †	Δ_2^P †	Δ_2^P †
$\subseteq_P$-optimal	in P	coNP	coNP	NP
Pareto-optimal	in P	coNP	coNP	NP
Completion-optimal	in P	coNP	coNP	NP
Globally-optimal	coNP	Π_2^P	Π_2^P	Σ_2^P
$\Pi_{\subseteq}$-preferred	coNP	Π_2^P	Π_2^P	Σ_2^P
$\Pi_{\leq}$-preferred	coNP	Θ_2^P	Θ_2^P	Θ_2^P

†: Θ_2^P if there is a data-independent bound on the weights/number of priority levels

Upper bounds: $\subseteq$-repairs and languages with consistency checking/BCQ entailment in P

Lower bounds: DL-Lite$_{core}$, functional dependencies, or negative constraints

Three cases:

- Repair checking/CQA have same complexity as with standard repairs: "local" preference, no need to guess/compute another preferred repair
- Can compute the value of a global parameter (weight...) and use it to check that a $\mathcal{T}$-consistent subset of $\mathcal{D}$ is a preferred repair
- No better option than relying on the naïve guess-and-check algorithm to decide whether a repair is (not) preferred

Implementations of (preferred) repair-based semantics

Existing implementations

- Target different settings: database with different kinds of constraints and queries, knowledge bases in different languages
- Target different semantics: CQA, intersection, brave, others, based on standard or different kinds of preferred repairs
- Often rely on external solvers (SAT, ASP, BIP...) for hard problems
- Often use rewriting techniques for simpler problems

Implementations of (preferred) repair-based semantics

Example of SAT-based approach: the ORBITS system

Semantics considered by ORBITS:

- CQA, intersection and brave semantics
- Standard, Pareto- and completion-optimal repairs

Setting: prioritized KB (or database with denial constraints) $\mathcal{K}_{\succ}$

- Case where conflicts contain at most two facts: conflicts and priority relation can be represented as a directed graph such that there is an edge from α to β if $\{\alpha, \beta\}$ is a conflict and $\alpha \not\succ \beta$
- Input: directed conflict graph + potential answers and their causes (minimal sets of facts that support the answer)
- Output: answers that hold under the required semantics

[Bienvenu and Bourgaux, 2022]

High-level algorithm:

- Filter answers that trivially holds under (preferred repair-based) intersection semantics in polynomial time: those which have causes without any fact with outgoing edge in the directed conflict graph
- Check remaining potential answers using a SAT solver
  - possibility to choose among several algorithms and encoding variants

High-level ideas underlying SAT encodings:

- Try to build a subset of $\mathcal{D}$ that fulfills some conditions: assigning variable x_α to true means that fact α belongs to the subset
- Consider only relevant facts
- X-CQA: build a set of facts that can be extended to an X-optimal repair that does not contain any cause for the query
- X-brave: build a set of facts that contains a cause of the query and can be extended to an X-optimal repair
- X-intersection: for each cause, find a set of facts that does not contain it and can be extended to an X-optimal repair

Modular SAT encodings with basic building blocks:

- Absence of a cause (two encoding variants)
- Presence of a cause
- Consistency of selected facts
- Extension to X-optimal repair (two variants for Pareto-optimal repairs)

Example: CQA based on Pareto-optimal repairs

$$\Phi_{\text{P-CQA}}(q) = (\bigwedge_{\mathcal{C} \in \mathit{Causes}(q,\mathcal{K})} \varphi_{\neg\mathcal{C}}) \wedge \varphi_{\text{P-max}}(F) \wedge \varphi_{\text{cons}}(F') \text{ where:}$$

$$\varphi_{\neg\mathcal{C}} = \bigvee_{\alpha \in \mathcal{C}} \bigvee_{\alpha \perp \beta, \alpha \not\succ \beta} x_\beta \qquad \varphi_{\text{P-max}}(F) = \bigwedge_{\alpha \in \mathsf{R}(F)} (x_\alpha \vee \bigvee_{\alpha \perp \beta, \alpha \not\succ \beta} x_\beta)$$

$$\varphi_{\text{cons}}(F') = \bigwedge_{\alpha, \beta \in F', \alpha \perp \beta} (\neg x_\alpha \vee \neg x_\beta)$$

$F = \{\beta \mid x_\beta$ occurs in $\bigwedge_{\mathcal{C} \in \mathit{Causes}(q,\mathcal{K})} \varphi_{\neg\mathcal{C}}\}$, $F' = \{\beta \mid x_\beta$ occurs in $\varphi_{\text{P-max}}(F)\}$, and $\mathsf{R}(F)$ is the set of facts reachable from F in the directed conflict graph

Implementations of (preferred) repair-based semantics

Example of SAT-based approach: the ORBITS system

Algorithms

- Four generic algorithms, applicable to X-CQA, X-brave and X-intersection
  - one makes a single SAT call for each candidate answer
  - the others treat all candidate answers together (global encoding with soft clauses representing answers) with different reasoning modes
- Another algorithm for X-brave and X-intersection
  - check cause by cause
- Two algorithms for X-intersection keeping track of facts in the intersection of X-optimal repairs
  - cause by cause and fact by fact
  - all relevant facts together

Encouraging experimental results but the choice of algorithm/encoding variant for a given semantics may make a significant difference

Runtimes of query answering under CQA with standard versus optimal repairs: depend of the specific problem

Conclusion

- Many inconsistency-tolerant semantics, most of them based on repairs
- Different kinds of preferred repairs
- Some repairs related to other formalisms for inconsistency-handling (abstract argumentation, active integrity constraints)
- Using preferred repairs often increases the computational complexity of reasoning but not always
- Implemented systems, often based on reductions and use of solvers

Outlook

Inconsistency-handling with (preferred) repairs is an active line of research

- Practical algorithms and implementations still lacking for many cases
- Extensions of repair-based semantics to new settings
  - RDF graphs and SHACL constraints
  - graph databases
  - temporal databases/KBs
  - ...
- Reasoning tasks beyond query answering
  - query result explanations (why/why not true under a given semantics)
  - abduction
  - ...

References

Arenas, M., Bertossi, L. E., and Chomicki, J. (1999).
Consistent query answers in inconsistent databases.
In *Proceedings of the 18th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)*.

Arming, S., Pichler, R., and Sallinger, E. (2016).
Complexity of repair checking and consistent query answering.
In *Proceedings of the 19th International Conference on Database Theory (ICDT)*.

Belabbes, S., Benferhat, S., and Chomicki, J. (2021).
Handling inconsistency in partially preordered ontologies: the Elect method.
Journal of Logic and Computation, 31(5):1356–1388.

Benferhat, S., Bouraoui, Z., and Tabia, K. (2015).
How to select one preferred assertional-based repair from inconsistent and prioritized DL-Lite knowledge bases?
In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

Bienvenu, M. (2020).
A short survey on inconsistency handling in ontology-mediated query answering.
Künstliche Intelligenz, 34(4):443–451.

Bienvenu, M. and Bourgaux, C. (2020).
Querying and repairing inconsistent prioritized knowledge bases: Complexity analysis and links with abstract argumentation.
In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

Bienvenu, M. and Bourgaux, C. (2022).
Querying inconsistent prioritized data with ORBITS: Algorithms, implementation, and experiments.
In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

Bienvenu, M. and Bourgaux, C. (2023).
Inconsistency handling in prioritized databases with universal constraints: Complexity analysis and links with active integrity constraints.
In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

Bienvenu, M., Bourgaux, C., and Goasdoué, F. (2014).
Querying inconsistent description logic knowledge bases under preferred repair semantics.
In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*.

Bienvenu, M., Bourgaux, C., and Jean, R. (2025).
A rule-based approach to specifying preferences over conflicting facts and querying inconsistent knowledge bases.
In *Proceedings of the 22nd International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

Bienvenu, M. and Rosati, R. (2013).
Tractable approximations of consistent query answering for robust ontology-based data access.
In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*.

Calautti, M., Greco, S., Molinaro, C., and Trubitsyna, I. (2022).
Preference-based inconsistency-tolerant query answering under existential rules.
Artificial Intelligence (AIJ), 312:103772.

Du, J., Qi, G., and Shen, Y. (2013).
Weight-based consistent query answering over inconsistent $\mathcal{SHIQ}$ knowledge bases.
Knowledge and Information Systems, 34(2):335–371.

Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., and Savo, D. F. (2010).
Inconsistency-tolerant semantics for description logics.
In *Proceedings of the 4th International Conference on Web Reasoning and Rule Systems (RR)*.

Lopatenko, A. and Bertossi, L. E. (2007).
Complexity of consistent query answering in databases under cardinality-based and incremental repair semantics.
In *Proceedings of the 11th International Conference on Database Theory (ICDT)*.

Mahmood, Y., Virtema, J., Barlag, T., and Ngomo, A. N. (2024).
Computing repairs under functional and inclusion dependencies via argumentation.
In *Proceedings of the 13th International Symposium on Foundations of Information and Knowledge Systems (FoIKS)*.

Staworko, S., Chomicki, J., and Marcinkowski, J. (2012).
Prioritized repairing and consistent query answering in relational databases.
Annals of Mathematics and Artificial Intelligence (AMAI), 64(2-3):209–246.